GWAS in the SIGNAL/PHARE clinical cohort restricts the association between the FGFR2 locus and estrogen receptor status to HER2-negative breast cancer patients

David G Cox, Elsa Curtit, Gilles G Romieu, Pierre G Fumoleau, Maria Rios, Hervé Bonnefoi, Thomas Bachelot, Patrick Soulié, Christelle G Jouannaud, Hugues Bourgeois, et al.

▶ **To cite this version:**

David G Cox, Elsa Curtit, Gilles G Romieu, Pierre G Fumoleau, Maria Rios, et al.. GWAS in the SIGNAL/PHARE clinical cohort restricts the association between the FGFR2 locus and estrogen receptor status to HER2-negative breast cancer patients . Oncotarget, 2016, 10.18632/oncotarget.12669 . hal-01391480

HAL Id: hal-01391480
https://hal.science/hal-01391480

Submitted on 8 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GWAS in the SIGNAL/PHARE clinical cohort restricts the association between the *FGFR2* locus and estrogen receptor status to HER2-negative breast cancer patients

Author List

David G. Cox,[1*] Elsa Curtit,[2] Gilles Romieu,[3] Pierre Fumoleau,[4] Maria Rios,[5] Hervé Bonnefoi,[6] Thomas Bachelot,[7] Patrick Soulié,[8] Christelle Jouannaud,[9] Hugues Bourgeois,[10]Thierry Petit,[11] Isabelle Tennevet,[12] David Assouline,[13] Marie-Christine Mathieu,[14] Jean-Philippe Jacquin,[15] Sandrine Lavau-Denes,[16] Ariane Darut-Jouve,[17] Jean-Marc Ferrero,[18] Carole Tarpin,[19] Christelle Lévy,[20] Valérie Delecroix,[21] Véronique Trillet-Lenoir,[22]Oana Cojocarasu,[23] Jérôme Meunier,[24] Jean-Yves Pierga,[25] Céline Faure-Mercier,[26] Hélène Blanché,[27]Mourad Sahbatou,[27] Anne Boland,[28] Delphine Bacq,[28] Céline Besse,[28]Jean-François Deleuze,[27,28] Iris Pauporté,[26]Gilles Thomas,[29] Xavier Pivot[2]

Affiliations

[1]Centre de Recherche en Cancérologie de Lyon, INSERM U1052 - Centre Léon Bérard, 28 rue Laennec, 69373 Lyon, France

[2]Hôpital Jean-Minjoz, Centre Hospitalier Universitaire, Boulevard Fleming, 25030 Besançon, France.

[3]Oncologie Sénologie, ICM Institut Régional du Cancer, 34298 Montpellier Cedex, France

[4]Georges-François Leclerc, 1 Rue du Professeur Marion, 21000 Dijon, France

[5]Institut de Cancérologie de Lorraine - Alexis Vautrin, département d'Oncologie Médicale, 6, avenue de Bourgogne, 54511 VANDOEUVRE LES NANCY cedex, France

[6]Institut Bergonié, Département d'Oncologie Médicale, 229 Cours de l'Argonne, 33000 Bordeaux, France

[7]Centre Léon Bérard, Département de Cancérologie Médicale, 28 rue Laënnec, Lyon Cedex 08, France

[8]Institut de Cancérologie de l'Ouest, Service Oncologie Médicale, 2 rue Moll, 49993 Angers Cedex 09, France

[9]Institut Jean Godinot, Service Oncologie Médicale, 1 rue du Général Koenig, 51056 Reims cedex, France

[10]Clinique Victor Hugo-Centre Jean Bernard, 18 rue Victor Hugo, 72015 Le Mans cedex 2, France

[11]Centre Paul Strauss, Service d'Oncologie Médicale, 3 rue de la Porte de l'Hôpital, 67065 Strasbourg cedex, France

[12]Centre Henri Becquerel, rue d'Amiens, 76038 Rouen, France

[13]Institut Daniel Hollard, Service Oncologie Médicale, 8 rue du Docteur Calmette, 38028 Grenoble cedex 01, France

[14]Institut Gustave Roussy, Comité de Pathologie mammaire, 39 rue Camille Desmoulins, 94805 Villejuif cedex, France

[15]Institut de Cancérologie Lucien Neuwirth, Service Oncologie Médicale, 108 bis avenue Albert Raimond, 42270 Saint Priest en Jarez,France

[16]Centre Hospitalier de Limoges, Service d'Oncologie Médicale, 2 avenue Martin Luther King, 87042Limoges cedex, France

[17]Clinique Drévon, Centre d'oncologie et de radiothérapie du Parc , 18 cours du général de Gaulle, 21000 Dijon, France

[18]Centre Antoine Lacassagne, Département Oncologie Médicale, 33 avenue de Valombrose, 06189 Nice cedex 02, France

[19]Institut Paoli-Calmettes, Département d'Oncologie Médicale, 232 Boulevard de Sainte-Marguerite, 13009 Marseille, France

[20]Centre François Baclesse, 3 avenue du Général Harris, 14076 Caen cedex 5, France

[21]Pôle Mutualiste, Service Oncologie Médicale, 11 boulevard Georges Charpak, 44606 Saint Nazaire, France

[22]Centre Hospitalier Lyon Sud, Service d'Oncologie Médicale, 165 chemin du Grand Revoyet, 69495 Pierre-Benite cedex, France

[23]Centre Hospitalier Le Mans, Service d'Onco-Hématologie et Médecine interne, 194 avenue Rubillard, 72037Le Mans Cedex, France

[24]Centre Hospitalier Régional d'Orléans, Service d'Oncologie médicale, 1 rue Porte Madeleine, 45032 ORLEANS Cedex 1, France

[25]Institut Curie, Department of MedicalOncology, 26 rue d'Ulm, 75248 Paris Cedex 05, France

[26]Institut National du Cancer, Direction de la Recherche, 52 avenue Morizet, 92513 Boulogne-Billancourt, France

[27]Fondation Jean Dausset – CEPH (Centre d'Etudes du Polymorphisme Humain), 27 rue Juliette Dodu, 75010 Paris, France

[28]Centre National du Génotypage, Institut de Génomique, CEA, 2 rue Gaston Crémieux, CP 5721, 91057 Evry Cedex, France

[29]Synergie Lyon Cancer, Centre Léon Bérard, 28 rue Laënnec, Lyon cedex 08, France.

* Correspondance to david.cox@inserm.fr

Abstract

Genetic polymorphisms are associated with breast cancer risk. Clinical and epidemiological observations suggest that clinical characteristics of breast cancer, such as estrogen receptor or HER2 status, are also influenced by hereditary factors. To identify genetic variants associated with pathological characteristics of breast cancer patients, a Genome Wide Association Study was performed in a cohort of 9365 women from the French nationwide SIGNAL/PHARE studies (NCT00381901/RECF1098). Strong association between the *FGFR2* locus and ER status of breast cancer patients was observed (ER-positive n=6211, ER-negative n=2516; rs3135718 OR=1.34 $p=5.46x10^{-12}$). This association was limited to patients with HER2-negative tumors (ER-positive n=4267, ER-negative n=1185; rs3135724 OR=1.85 $p=1.16x10^{-11}$). The *FGFR2* locus is known to be associated with breast cancer risk. This study provides sound evidence for an association between variants in the *FGFR2* locus and ER status among breast cancer patients, particularly among patients with HER2-negative disease. This refinement of the association between *FGFR2* variants and ER-status to HER2-negative disease provides novel insight to potential biological and clinical influence of genetic polymorphisms on breast tumors.

Introduction

Since the completion of the Human Genome Project, the Genome Wide Association Scan (GWAS) has become the tool of choice for the detection of associations between disease risk, and common genetic variation. The first breast cancer risk variants identified in the GWAS era were in the *FGFR2* locus [1,2].

Further analyses, mainly in case-control and prospective cohorts, have reinforced this association as well as identified over 90 additional breast cancer risk loci [3]. GWAS studies with cases selected based on the estrogen receptor (ER) status of their tumors, and control subjects not affected by breast cancer, have shown divergent associations between ER+ and ER- tumors. In these analyses, variants in *FGFR2*are more strongly associated with ER+ disease [4-14], as opposed to ER- disease, when comparing cases to healthy controls.Few single studies, however, have sufficient detail or sample size to carry out case-only analyses to further explore the relationship between genetic variants and disease characteristics, particularly with respect to amplification of the *HER2* gene.Therefore analyses by subtype are often secondary, based on findings of the primary analyses of overall breast cancer risk. Furthermore, these studies are now carried out in large consortia with the potential for heterogeneity in definitions of various case characteristics, particularly ER and HER2 status. For example, Broeks et al. [13] examined the association between low penetrance breast cancer loci and specific breast tumor subtypes in the context of the Breast Cancer Association Consortium (BCAC). rs2981582 in the *FGFR2* locus was significantly associated with ER+/PR+/HER2- breast cancer (n_{cases}=7201, p = 2.2 x 10^{-29}), less so with ER+/PR+/HER2+ cases (n_{cases}=996, p=5.5x10^{-4}), and no association was observed with triple negative breast cancer (n_{cases}=1480, p=0.841) or ER-/PR-/HER2+ breast cancer (n_{cases}=627, p=0.396). A case-only comparison of HER2 status was carried out within ER+/PR+ and ER-/PR- groups, and neither showed any association (p=0.23 and 0.15, respectively).

In the present study, a case-only GWAS approach was used to study differences in the distribution of variants between breast cancer cases in a large, multi-center study with centralized data collection and handling, the SIGNAL/PHARE case-cohorts (NCT00381901/RECF1098).

Results

Genotype data was generated from 9365 SIGNAL/PHARE participants. All subjects had greater than 95% genotyping success rate. 26 pairs of individuals were identified with Identity by State (IBS) > 30%, with the subject having the most complete genotype data from each pair retained for analyses. 551 further individuals were excluded from the present study due to PCA analyses. Finally, 61 subjects with missing clinical data were excluded. A total of 8727 patients including 2516 patients with ER- breast cancer were analyzed. Furthermore, 5452 patients had HER2-negative breast cancer, of which 1185 were ER-.

The search for variants associated with ER status showed only one region with a highly significant association, corresponding to *FGFR2* (best p-value for rs3135718 p-value=6.0×10^{-12}, Figure 1). Restricting our analyses to HER2-negative cases found that associations between variants at the *FGFR2* locus remained significant at the genome-wide level (best p-value for rs3135724 = 5.2×10^{-11}, Figure 2). Among HER2-positive tumors, the lowest p-value in the *FGFR2* locus for the association with ER status was found for rs2981578 ($p = 3.3 \times 10^{-4}$ Table 1). The four variants in Table 1 were chosen to highlight the difference in associations between HER2+ and HER2- patients. Despite the smaller sample size among HER2-positive cases, this study has nearly 100% power to detect a per-allele OR =1.8 as observed among the HER2-negative tumors, and greater than 80% power to detect a per-allele OR ≈ 1.3. The observed direction of the association was consistent with observations in prior case-control studies, with for example the C allele of rs3135718 being more frequently reported among women with ER+ tumors.

As mentioned previously, variants in the *FGFR2* locus were the first identified via GWAS with respect to breast cancer risk. The most recent fine-mapping effort of the *FGFR2* locus explored functional variants, and identified three separate independent sets of correlated highly associated variants (ICHAVs [18]). In the present analyses restricted to HER2-negative tumors, rs3135724 was the SNP with the strongest association for ER status.

These data included rs2981579 and rs2981578, from ICHAVs 1 and 3 respectively (Table 1). Unfortunately, rs45631563 from ICHAV 2 was not included, and no SNPs showed significant linkage disequilibrium with this marker in the current 1000 genomes data (http://1000genomes.org accessed July 8, 2015). Therefore additional analyses were carried out including rs3135724, rs2981579, and rs2981578 in the same logistic regression model. In our analyses of HER2- breast cancer, we found no evidence for independent association between these variants and tumor ER status (data not shown).

Discussion

The identification of variants associated with specific molecular subtypes of breast cancer was a primary aim of the prospective SIGNAL/PHARE cohort. In this high-powered GWAS performed in a case-cohort of breast cancer patients with detailed clinical data, further information with respect to variants in the *FGFR2* locus and their influence on breast cancer were provided, particularly regarding tumor ER status. In addition, the association between variants in *FGFR2* and ER status in breast cancer was stronger among patients with HER2- tumors. While not including an independent validation set is a drawback of our analyses, the large sample size allowed us to have sufficient power to fully define this association, and the p-values obtained were well below empirical estimations of significance thresholds ($1.48x10^{-7}$) as well as the generic GWAS significance threshold of $5x10^{-8}$.

Our hypothesis is that genetic variants that are associated with molecular subtypes will provide novel insights regarding disease etiology, and may lead to further developments regarding disease prevention and treatment. As our main focus was the construction of a clinical cohort, we have focused on collecting information with respect to histo-pathology and treatments, and patient follow-up. Therefore, we have not collected detailed information regarding epidemiological data such as body-mass index, reproductive history and menopausal status, or family history/BRCA mutations. The participants have been given a self-administered questionnaire with some of these variables, but as this questionnaire was administered after cancer diagnosis, we have chosen to not exploit these data at this time.

We have focused on the *FGFR2* locus, which showed the strongest association with ER status, particularly among HER- breast cancer patients. There is growing evidence that genetic variants may be more strongly associated with specific breast cancer subtypes. For the most part, these analyses are extensions of current prospective cohort and case-control analyses. For example, recent analyses by Michailidou et al. [3] included stratification by estrogen receptor status for the 77 variants included in their polygenic risk score. A number

of these variants showed differential associations with respect to estrogen receptor status. However as the authors state in their discussion, the number of estrogen receptor negative cases made accurately determining risk estimates difficult for this cancer subtype.Future analyses in our case-cohort will investigate other variants previously shown to influence breast cancer subtype.

A potential limitation of our study is the use of an internal imputation process, as opposed to imputing to the commonly used 1000 Genomes data or the Michigan Imputation Server. As mentioned previously, this was our original study design prior to the availability of these resources. We have continued with this approach in order to avoid any potential population differences with respect to linkage disequilibrium between our population of French breast cancer cases and the populations that provided data for publicly available resources. This approach leads to a lower number of variants on the absolute scale, meaning that we may be unable to detect any additional variants not captured through genotyping with the Illumina Omni5, which captures over 80% of common variants among Caucasian populations, and strict quality filtering of data (See Methods section).

For aspects of response to treatment, SIGNAL/PHARE has not yet accrued enough follow-up to fully explore the implication of variants on patient's outcome. This will be of course an obvious next step of our analyses, particularly as pertains to response to hormone therapy and *FGFR2* variants in ER+/HER- breast cancer patients.

In conclusion, we further refine the influence of variants in the *FGFR2* locus with respect to molecular characteristics of breast tumors, in that they are more strongly associated with estrogen receptor status among cancers without amplification of the *HER2* gene.

Methods

PHARE was a randomized phase 3 clinical trial comparing 6- and 12-month trastuzumab adjuvant exposure [15], which included a subset of 1,430 HER2-positive breast cancer cases with DNA available for GWAS analyses. SIGNAL was a prospective cohort specifically designed for GWAS analyses of 8,406 early breast cancer patients, enrolled at the time of the adjuvant chemotherapy from June 2009 to December 2013. The combined data set, the PHARE/SIGNAL study, included 9,365 breast cancer patients. Clinical and pathological characteristics were prospectively collected using standardized forms, and centralized at the French National Cancer Institute (INCa). For both studies, patients provided blood samples that were centralized at the Centre d'Etude du PolymorphismeHumain (CEPH) in Paris, France, for DNA extraction using standard protocols. Genotyping was carried out at the Centre National de Génotypage (CNG) in Evry, France.

The original study plan called for a two-staged genotyping strategy using only study participants. This approach aimed at reducing the potential that population structure in French breast cancer cases would influence imputation, while maximizing the proportion of the genome covered. Briefly, all cases were genotyped using the IlluminaHumanCoreExome chip set, composed of over 264000 variants for a "GWAS Backbone" and over 244000 "exome-centered" variants. Variants were filtered based on completion rates (<95% SNP success, N = 8122), departure from Hardy-Weinberg Equilibrium (HWE $p<0.001$, N = 20357), and low minor allele frequency (MAF<0.001, N=200628). Principal Components Analysis (PCA) and k-means were then used to characterize the ancestry of the participants and only the main cluster of European individuals was included in the present analysis, to reduce risk of population stratification (See Figure S1). A random subset of 1449 individuals from the main "European" cluster was selected for genotyping using the Illumina Omni5 chip set, composed of over 4M variants (See Figure S1). Complete (SNP success = 100%, N=2049173) Omni5 data were then filtered using similar cutoffs as the HumanCoreExome data, specifically HWE ($p<0.001$, N=91018) were then used to impute

missing genotypes from the remaining subjects genotyped using the HumanCoreExome array. SNPs with imputation quality score < 30% were excluded from analyses (N=783416), and finally variants with a MAF < 0.01 were excluded (N=82847). A total of 914144 SNPs were included in the GWAS analyses. Standard GWAS logistic regressions were carried out using the ProbABEL package [16]. Age at diagnosis and the first two principal components were included in regression analyses.

Genome-wide significance levels were estimated using the effective number of tests based on linkage disequilibrium between all markers used in our population through the SimpleM function in R [17]. The number of effective markers is estimated at 345906, corresponding to a Bonferroni-corrected p-value threshold of 1.48×10^{-7}.

Acknowledgements

The authors would like to thank the patients who took part in the study as well as medical staff for their cooperation. We also would like to thank Nicolas Thammavong for his contribution in clinical data management, Alexia Renoud for her contribution in clinical data management and descriptive statistics analyses, and the CEPH Biological Resource Center and CNG genotyping staff for technical assistance.

Conflict of Interests

The authors have no conflicts of interest to declare.

Grant Support

PHARE and SIGNAL are academic trials sponsored by the French National Cancer Institute (INCa). David G. Cox receives support from the LigueContre le Cancer, Comité de l'Ain and the association Amis de l'Université de Lyon 1

References

1. Hunter, D.J. Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Want, Z., Welch, R., Hutchinson, A. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.***39,** 870–874 (2007).
2. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Stuewing J.P., Morrison J., Field, H., Luben, R. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature***447,** 1087–1093 (2007).
3. Michailidou, K., Beesley, J., Lindstrom S., Canisius, S., Dennis, J., Lush, M.J., Maranian, M.J., Bolla, M.K., Wang, Q., Shah, M., *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.***47,** 373–380 (2015).
4. Siddiq, A., Couch, F.J., Chen, G.K., Lindström, S., Eccles, D., Millikan, R.C., Michailidou, K., Stram, D.O., Beckmann, L., Rhie, K.S, *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum. Mol. Genet.***21,** 5373–5384 (2012).
5. Haiman, C.A., Chen, G.K., Vachon C.M., Canzian F., Dunning, A., Millikan, R.C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C.B., *et al.* A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.***43,** 1210– 1214 (2011).
6. Figueroa, J.D., Garcia-Closas, M., Humphries, M., Platte, R., Hopper, J.L., Southey, M.C., Apicella C., Hammet, F., Schmidt, M.K., Broeks, A., *et al.* Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Hum. Mol. Genet.***20,** 4693–4706 (2011).

7. Stevens, K.N., Fredericksen, Z., Vachon, C.M., Wang, X., Margolin, S., Lindblom, A., Nevanlinna, H., Breco D., Aittomäki, K., Blomqvist, C., *et al.* 19p13.1 is a triple-negativespecific breast cancer susceptibility locus. *Cancer Res.***72,** 1795–1803 (2012).

8. Garcia-Closas, M., Couch F.J., Lindstrom, S., Michailidou, K., Schmidt, M.,K., Brook, M.N., Orr,

   N., Rhie, S.K., Riboli, E., Feigelson, H.S., *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.***45,** 392–398, 398e1–2 (2013).

9. Lambrechts, D., Truong, T., Justenhoven, C., Humphreys, M.K., Wang, J., Hopper, J.L., Dite, G.S., Apicella, C., Southey, M.C., Schmidt, M.K., *et al.* 11q13 is a susceptibility locus for hormone receptor positive breast cancer. *Hum. Mutat.***33,** 1123–1132 (2012).

10. Purrington, K.S., Slager, S., Eccles, D., Yannoukakos, D., Fasching, P.A., Miron, P., Carpenter, J., Chang-Claude, J., Martin, N.G., Montgomery, G.W., *et al.* Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis***35,** 1012–1019 (2014).

11. Warren, H., Dudbridge, F., Fletcher, O., Orr, N., Johnson, N., Hopper, J.L., Apicella, C., Southey, M.C., Mahmoodi, M., Schmidt, M.K., *et al.* 9q31.2-rs865686 as a susceptibility locus for estrogen receptor-positive breast cancer: evidence from the Breast Cancer Association Consortium. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.***21,** 1783–1791 (2012).

12. Stevens, K.N., Vachon, C.M., Lee, A.M., Slager, S., Lesnick, T., Olswold, C., Fasching, P.A., Miron, P., Eccles, D., Carpenter, J.E., *et al.* Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res.***71,** 6240–6249 (2011).

13. Broeks, A., Schmidt, M.K., Sherman, M.E., Couch, F.J., Hopper, H.L., Dite, G.S., Apicella, C., Smith, L.D., Hammet, F., Southey, M.C., *et al.* Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum. Mol. Genet.***20,** 3289–3303 (2011).

14. Cen, Y.-L., Qi, M.-L., Li, H.-G., Su, Y., Chen, L.-J., Lin, Y., Chen, W.-Q., Xie, X.-M., Tang, L.-Y., and Ren, Z.-F. Associations of polymorphisms in the genes of FGFR2, FGF1, and RBFOX2 with breast cancer risk by estrogen/progesterone receptor status. *Mol. Carcinog.***52 Suppl 1,**E52–59 (2013).

15. Pivot, X. Romieu, G., Debled, M., Pierga, J.-Y., Kerbrat, P., Bachelot, T., Lortholary, A., Espié, A., Fumoleau, P. Serin, D., *et al.* 6 months versus 12 months of adjuvant trastuzumab for patients with HER2-positive early breast cancer (PHARE): a randomised phase 3 trial. *Lancet Oncol.***14,** 741–748 (2013).

16. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*. **11**:134 (2010)

17. Goa, X., Becker, L.C., Becker, D.M., Starmer, J.D. & Province, M.A. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.***34**, 100-105 (2010)

18. Meyer, K.B., O'Reilly, M., Michailidou, K., Saskia, C., Edwards, S.L., French, J.D., Prathalingham, R., Dennis, J., Bolla, M.K., Wang, Q., *et al.* Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am. J. Hum. Genet.***93,** 1046–1060 (2013).

Table 1. Selected variants at the *FGFR2* locus and ER status among breast cancer cases

		Overall		HER2+		HER2-	
SNP	I/G* (Rsq, Quality)	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
rs3135718	I (0.64, 0.89)	1.33 (1.23 - 1.45	6.0x10-12	1.19 (1.04 - 1.35)	7.9x10-3	1.47 (1.30 - 1.64)	2.0x10-10
rs3135724	I (0.41, 0.84)	1.51 (1.33 - 1.69)	8.1x10-11	1.18 (0.97 - 1.41)	9.3x10-2	1.79 (1.49 - 2.13)	5.2x10-11
rs2981578	G (NA, NA)	1.24 (1.16-1.32)	3.5x10-10	1.20 (1.09-1.33)	3.3x10-4	1.26 (1.14-1.38)	1.7x10-6
rs2981579	G (NA, NA)	1.25 (1.16 - 1.33)	5.5x10-11	1.15 (1.03 - 1.27)	9.2x10-3	1.33 (1.20 - 1.47)	2.1x10-9

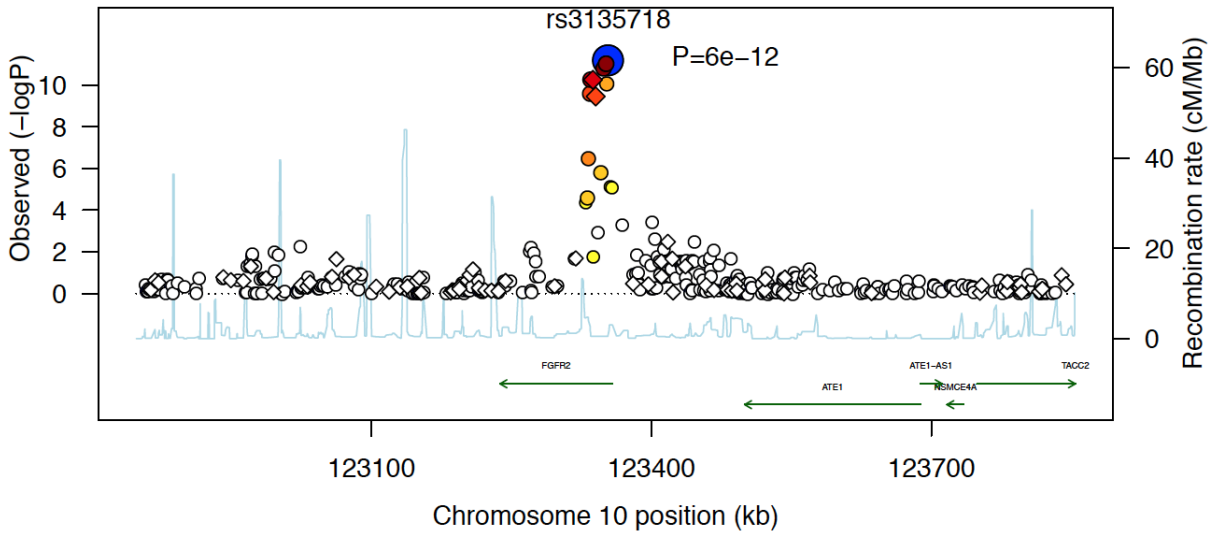
* Imputed (I) or genotyped (G). Values reported from MACH output

Figure 1. Manhattan plot of associations between SNPs and ER status overall.

P-values from logistic regression comparing estrogen receptor positive cases to estrogen receptor negative cases, controlling for age at diagnosis and first two principal components, are shown. rs3135718 on chromosome 10 at the *FGFR2* locus shows the strongest association. 914144 SNPs were included in these analyses, with 6211 ER+ and 2516 ER- cases. The red horizontal line corresponds to the empirical significance threshold of 1.48×10^{-7}, while the blue horizontal line corresponds to an arbitrary level of 1.0×10^{-5}. The inflation factor (λ) for these analyses is 1.02.

Figure 2. Manhattan plot of associations between SNPs and ER status restricted to HER2- cases.

P-values from logistic regression comparing estrogen receptor positive cases to estrogen receptor negative cases restricted to HER2- cases, controlling for age at diagnosis and first two principal components, are shown. rs2981578 on chromosome 10 at the *FGFR2* locus shows the strongest association. The same 914144 SNPs were included in these analyses, with 4267 HER2-/ER+ and 1185 HER2-/ER- cases. The red horizontal line corresponds to the empirical significance threshold of 1.48×10^{-7}, while the blue horizontal line corresponds to an arbitrary level of 1.0×10^{-5}. The inflation factor (λ) for these analyses is 1.02.



Supplementary Figure 1: Locuszoom plot of chromosome 10 around the FGFR2 locus. Circles represent imputed SNPs, diamonds represent genotyped SNPs. Observed p-value is plotted along the left Y axis, recombination rate along the right Y axis. Shading from purple to yellow in filled shapes represents linkage disequilibrium with the highlighted SNP, in this case rs3135718.